OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.3

Run on: January 24, 2003, 11.42.22 ; Search time 66.3129 Seconds
                 (without alignments)
                 11410.644 Million cell updates/sec

Title: us-09-910-428-2
Perfect score: 26
Sequence: 1 cctccccaat cgttttgaag t 26

Scoring table: IDENTITY_NUC
                 GapOp 10.0 GapExt 1.0

Searched: 2054640 seqs 1451402974 residues

Total number of hits satisfying chosen parameters: 4199280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl.*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mus:*
                 20: em_om:*
                 21: em_ov:*
                 22: em_pat:*
                 23: em_ph:*
                 24: em_pl:*
                 25: em_pr:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_plo:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_htg_vrt:*
                 38: em_syt:*
                 39: em_htgo_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1	c	26	100.0	732	4	AF040955	AF040955 Bos indic
2	c	26	100.0	25688	4	U15731	U15731 Bos tauru
3	c	22.8	87.7	349	4	AF126288	AF126288 Bos tauru
4	c	21.4	82.3	14569	2	AC125465	AC125465 Mus muscu
5	c	21.4	82.3	9461	9	AC020937	AC020937 Homo sapi
6	c	21.4	82.3	12345	2	AC009592	AC009592 Homo sapi
7	c	21.4	82.3	164478	2	AC008787	AC008787 Homo sapi
8	c	21.4	82.3	185217	2	AC124754	AC124754 Mus muscu
9	c	21.4	82.3	187592	2	AC069033	AC069033 Rattus no
10	c	21.2	81.5	186552	2	AC123369	AC123369 Mus muscu
11	c	21.2	81.5	223987	2	AC110785	AC110785 Homo sapi
12	c	20.2	77.7	108000	8	AF004970	AF004970 Lotus jap
13	c	20.2	77.7	157623	2	AC073450	AC073450 Homo sapi
14	c	20.2	77.7	230419	2	AC107804	AC107804 Mus muscu
15	c	19.8	76.2	95445	2	AC016175	AC016175 Homo sapi
16	c	19.8	76.2	99601	2	AL160492	AL160492 Homo sapi
17	c	19.8	76.2	99900	2	AC092518	AC092518 Homo sapi
18	c	19.8	76.2	124839	9	AC097109	AC097109 Homo sapi
19	c	19.8	76.2	127917	9	HS257A7	AL008729 Human DNA
20	c	19.8	76.2	147945	2	AC084735	AC084735 Homo sapi
21	c	19.8	76.2	163696	2	AC108616	AC108616 Rattus no
22	c	19.8	76.2	170776	9	AL139087	AL139087 Human DNA
23	c	19.8	76.2	174129	2	AC124533	AC124533 Mus muscu
24	c	19.8	76.2	183503	9	AL160253	AL160253 Human DNA
25	c	19.8	76.2	188488	2	AC125197	AC125197 Mus muscu
26	c	19.8	76.2	195145	2	AC119779	AC119779 Homo sapi
27	c	19.8	76.2	202553	2	AC124241	AC124241 Mus muscu
28	c	19.8	76.2	204764	2	AC125178	AC125178 Mus muscu
29	c	19.8	76.2	211342	2	AC125068	AC125068 Mus muscu
30	c	19.8	76.2	217320	2	AC124589	AC124589 Mus muscu
31	c	19.8	76.2	217935	2	AC121575	AC121575 Mus muscu
32	c	19.8	76.2	249698	2	AC124416	AC124416 Mus muscu
33	c	19.8	76.2	255787	2	AC140832	AC140832 Mus muscu
34	c	19.5	75.4	104424	9	AC095918	AC095918 Rattus no
35	c	19.5	75.4	115043	9	AL663070	AL663070 Human DNA
36	c	19.5	75.4	121156	9	AC127075	AC127075 Rattus no
37		19.5	75.4	136159	9	AC097475	AC097475 Homo sapi
38		19.5	75.4	136719	2	AC046436	AC046436 Homo sapi
39		19.5	75.4	136730	9	AC108110	AC108110 Homo sapi
40	c	19.5	75.4	154399	2	AC127822	AC127822 Rattus no
41	c	19.5	75.4	161000	2	AC097910	AC097910 Rattus no
42	c	19.5	75.4	163549	2	AC121648	AC121648 Rattus no
43	c	19.5	75.4	191191	2	AC115524	AC115524 Rattus no
44	c	19.5	75.4	195430	2	AC026347	AC026347 Homo sapi
45	c	19.5	75.4	200484	2	AC025002	AC025002 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS   AF040955   732 bp   DNA   linear   MAM 16-JAN-1998
DEFINITION   Bos indicus growth hormone receptor promoter and exon 1, partial
             sequence.
ACCESSION   AF040955
VERSION   AF040955.1 GI:2773337
KEYWORDS   .
SOURCE   Bos indicus
  ORGANISM   Bos indicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 732)
  AUTHORS   Lucy,M.C. and Boyd,C.K.

TITLE   Promoter and first exon for bovine growth hormone receptor isolated
          from Bos indicus
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 742)
AUTHORS   Lucy,M.C. and Boyd,C.K.
TITLE   Direct Submission
JOURNAL   Submitted (02-JAN 1998) Animal Science, University of Missouri, 164
          ASRC, Columbia, MO 65211, USA
FEATURES
     source
          1..742
          /organism="Bos indicus"
          /strain="Brahman"
          /db_xref="taxon:9915"
          /chromosome="20"
     promoter
          1..569
          /note="growth hormone receptor promoter"
     mRNA
          570..742
          /product="growth hormone receptor"
     exon
          570..742
          /number=1
BASE COUNT   176 a   168 c   175 g   213 t
ORIGIN

RESULT 2
BGU15741/c
LOCUS
DEFINITION   25688 bp   DNA   linear   MAM 04-SEP-2001
ACCESSION U15741
VERSION U15741.2   GI:14570041
KEYWORDS
SOURCE   Bos taurus
ORGANISM   Bos taurus
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
          Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 25688)
AUTHORS   Heap,D., Lucy,M.C., Collier,R.J., Boyd,C.K. and Warren,W.C.
TITLE   Rapid communication: nucleotide sequence of the promoter and first
          exon of the somatotropin receptor gene in cattle
JOURNAL   J. Anim. Sci. 74 (5), 1529 (1995)
PUBMED   7665485
REFERENCE 2 (bases 1 to 25688)
AUTHORS   Lucy,M.C.
TITLE   Direct Submission
JOURNAL   Submitted (11-OCT 1994) 164 Animal Science Research Center, University
          of Missouri, Columbia, MO
REFERENCE 3 (bases 1 to 25688)
AUTHORS   Jiang,H. and Lucy,M.C.
TITLE   Involvement of hepatocyte nuclear factor 4 in the expression of the
          growth hormone receptor 1A messenger ribonucleic acid in bovine
          liver
JOURNAL   Mol. Endocrinol. 15 (6), 1023-1034 (2001)
PUBMED   11376119
COMMENT   On Apr 10, 2001 this sequence version replaced gi:560024.
FEATURES
     source
          1..25688
          /organism="Bos taurus"
          /db_xref="taxon:9913"
          /clone="13-1"
          /sex="male"

RESULT 3
AF126288/c
LOCUS   AF126288   349 bp   DNA   linear   MAM 27-MAY 1999
DEFINITION   Bos taurus growth hormone receptor (GHR) gene, promoter and 5'
          untranslated region.
ACCESSION AF126288
VERSION   AF126288.1   GI:4894683
KEYWORDS
SOURCE   Bos taurus
ORGANISM   Bos taurus
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
          Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 349)
AUTHORS   Ge,W., Davis,M.E., Hines,H.C. and Irvin,K.M.
TITLE   Two-allelic PCR polymorphism detected in the promoter region of
          the bovine GHR gene
JOURNAL   Anim. Genet. 30 (1), 71 (1999)
PUBMED   10050296
REFERENCE 2 (bases 1 to 349)
AUTHORS   Ge,W., Davis,M.E., Hines,H.C. and Irvin,K.M.
TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-1999) Animal Sciences, the Ohio State University,
          2027 Coffey Road, Columbus, OH 43210, USA
FEATURES
     source
          1..349
          /organism="Bos taurus"
          /db_xref="taxon:9913"
          /chromosome="20"
     gene
          <1..>349
          /gene="GHR"
     promoter
          <1..302
          /gene="GHR"
     variation
          149
          /gene="GHR"
          /replace="4"
     mRNA
          303..>349
          /gene="GHR"

```
exon              /gene="GHR"
                  /product="growth hormone receptor"
                  303. .>349
                  /gene="GHR"
                  /number=1
5'UTR             303. .>349
                  /gene="GHR"

BASE COUNT     77 a     72 c    106 g     94 t
ORIGIN

Qy    1 CCTCCCCAAATCAATTACATTTCTC 25
        ||||  |||||||||||||||||||
Db  259 CCTCGGCAAATCAATTACATTTCTC 234

Query Match            87.8%;   Score 22.8;  DB 4;  Length 349;
Best Local Similarity  92.8%;   Pred. No. 50;
Matches   24   Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

RESULT 4
AC125465/c
LOCUS       AC125465            14568 bp    DNA    linear   HTG 26-JUN-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-333A16, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC125465
VERSION     AC125465.1  GI:21592226
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 14568)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       The sequence of Mus musculus clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 14568)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (26-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            --------------- Project Information ---------------
            Center project name: M_BA0333A16
            --------------- Summary Statistics ---------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: phrap; version 0.990319
            Consensus quality: 7558 bases at least Q40
            Consensus quality: 9287 bases at least Q30
            Consensus quality: 11208 bases at least Q20
            Insert size: 202000; agarose-fp
            Insert size: 13768; sum-of-contigs
            Quality coverage: 0.09 in Q20 bases; agarose-fp
            Quality coverage: 1.16 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *    1   1015; contig of 1015 bp in length
            * 1016   1115; gap of unknown length

FEATURES             Location/Qualifiers
     source          1..14568
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-333A16"
     misc_feature    1..1015
                     /note="assembly_name:Contig25"
     misc_feature    1116..2365
                     /note="assembly_name:Contig33"
     misc_feature    2466..3802
                     /note="assembly_name:Contig35"
     misc_feature    3903..5693
                     /note="assembly_name:Contig38"
     misc_feature    5794..7175
                     /note="assembly_name:Contig39"
     misc_feature    7276..8375
                     /note="assembly_name:Contig40"
     misc_feature    8476..9938
                     /note="assembly_name:Contig42"
     misc_feature    10039..11427
                     /note="assembly_name:Contig43"
     misc_feature    11528..14568
                     /note="assembly_name:Contig44"

     *   1116   2365; contig of 1250 bp in length
     *   2366   2465; gap of unknown length
     *   2466   3802; contig of 1337 bp in length
     *   3903   5693; contig of 1791 bp in length
     *   5694   5793; gap of unknown length
     *   5794   7175; contig of 1382 bp in length
     *   7276   8375; contig of 1100 bp in length
     *   8476   9938; contig of 1463 bp in length
     *  10039  11427; contig of 1389 bp in length
     *  11528  14568; contig of 3041 bp in length

Qy    3 TCCCCAAATCAATTACATTTTCT 25
        || |||||||||||||| |||||
Db 10186 TCCCCAAATCAATTACATATTCT 10164

BASE COUNT   4173 a   2750 c   2822 g   4022 t   801 others
ORIGIN

Query Match            82.3%;   Score 21.4;  DB 2;  Length 14568;
Best Local Similarity  95.7%;   Pred. No. 1.1e-02;
Matches   22   Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

RESULT 5
AC020937/c
LOCUS       AC020937            94161 bp    DNA    linear   HTG 21-SEP-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2239121, Complete sequence.
ACCESSION   AC020937
VERSION     AC020937.6  GI:15718535
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94161)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 94161)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 94161)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
```

Left column

TITLE   Direct Submission
JOURNAL   Submitted (30-AUG-2001) DoE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 94161)
AUTHORS   DoE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (21 SEP-2001) DoE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT   On Sep 21, 2001 this sequence version replaced gi:15375144.
           Draft Sequence produced by DoE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center
           www.shgc.stanford.edu
           Quality: Phrap Quality ~40 100% of Sequence;
           SFS Contig:
           Estimated Total Number of Errors is 0.
FEATURES
   source   1..94161
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="5"
BASE COUNT   28997 a   19265 c   18272 g   27627 t
ORIGIN
           W1 1219 G82845.

Qy   4 CCTCYAAATCYAATTACATTTTCTC 26
          |||||||||||||||||||||||||
Db 77874 CCCCAAATCCAATTAGATTTTCTC 77852

Query Match   82.38;   Score 21.4;   DB 9; Length 94161;
Best Local Similarity   95.78;   Pred. No. 87;
Matches 22;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

LOCUS   AC008502   124495 bp   DNA   linear   PRI 20-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTC 1123:11, complete sequence.
ACCESSION   AC008502
VERSION   AC008502.8 GI:14971159
KEYWORDS   HTG.
SOURCE   Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 124495)
AUTHORS   DoE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 124495)
AUTHORS   DoE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (01-JUN-2001) DoE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 124495)
AUTHORS   DoE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (20-JUL-2001) DoE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT   Draft Sequence Produced by DoE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality ~40 99.5% of Sequence;
            SFS Contig;
            Estimated Total Number of Errors is 0.4.

Right column

SHGC-33971 G29589.
FEATURES
   source   1..124495
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="5"

Qy   4 CCCCAAATCAATTACATTTTCTC 26
          |||||||||||||||||||||||
Db 8179 CCCCAAATCAATTAGATTTTCTC 8157

Query Match   82.38;   Score 21.4;   DB 9; Length 124495;
Best Local Similarity   95.78;   Pred. No. 84,
Matches 22;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

RESULT 7
ACCESSION   AC008787
LOCUS   AC008787/c   164478 bp   DNA   linear   HTG 05-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone CTD 2018E6, WORKING DRAFT SEQUENCE,
            29 unordered pieces.
ACCESSION   AC008787
VERSION   AC008787.4 GI:7609311
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164478)
AUTHORS   DoE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production sequencing facility, DoE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On May 5, 2000 this sequence version replaced gi:6694425.
            ----- Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 64224
            Center clone name: CITB-H1-2018E6

            Summary Statistics
            Consensus quality: 129975 bases at least Q40
            Consensus quality: 148252 bases at least Q30
            Consensus quality: 154251 bases at least Q20
            Estimated insert size: 162578; pulse-field gel estimation
            Estimated insert size: 162578; sum of contigs estimation
            Quality coverage: 5.12 in Q20 bases, pulse-field gel estimation
            Quality coverage: 4.95 in Q20 bases; sum of contigs estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be retained.

Summary Statistics
   1   226.2 contig of 226.2 bp in length
   2263   236.2 gap of unknown length
   2364   546.9 contig of 3107 bp in length
   5470   569.0 gap of unknown length
   5570   864.2 contig of 8073 bp in length
   8643   874.3 gap of unknown length
   8743   12920 contig of 4178 bp in length

```
*      12921      13020: gap of unknown length
*      13021      13620: contig of 3600 bp in length
*      16621      16720: gap of unknown length
*      16721      18847: contig of 2127 bp in length
*      18848      18947: gap of unknown length
*      18948      22210: contig of 3263 bp in length
*      22211      22310: gap of unknown length
*      22311      27037: contig of 4727 bp in length
*      27038      27137: gap of unknown length
*      27138      31662: contig of 4525 bp in length
*      31663      31762: gap of unknown length
*      31763      37223: contig of 5461 bp in length
*      37224      37323: gap of unknown length
*      37324      42764: contig of 5441 bp in length
*      42765      42864: gap of unknown length
*      42865      48849: contig of 5985 bp in length
*      48850      48949: gap of unknown length
*      48950      54895: contig of 5946 bp in length
*      54896      54995: gap of unknown length
*      54996      63708: contig of 8713 bp in length
*      63709      63808: gap of unknown length
*      63809      74072: contig of 10264 bp in length
*      74073      74172: gap of unknown length
*      74173      87006: contig of 12834 bp in length
*      87007      87106: gap of unknown length
*      87107     101079: contig of 13973 bp in length
*     101080     101179: gap of unknown length
*     101180     113781: contig of 12602 bp in length
*     113782     113881: gap of unknown length
*     113882     129485: contig of 15604 bp in length
*     129486     129585: gap of unknown length
*     129586     164478: contig of 34893 bp in length
       Location/Qualifiers
       1..164478
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="5"
       /clone_lib="CalTech human BAC library D"
       /clone="CTD-2038E6"

BASE COUNT    48924 a    32821 c    32760 g    48016 t    1937 others

ORIGIN

RESULT 8

LOCUS       AC124754             186217 bp    DNA    linear   HTG 15 JUN 2002
DEFINITION  Mus musculus chromosome UNK clone RP23-285C16, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC124754
VERSION     AC124754.1  GI:21428020
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 186217)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186217)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
```

```
Qy          4  CCCTAAATCAATTACATTCTC 26
               |||||||||||||||||||||
Db      84470  CCCTAAATCAATTACATTTTTC 84448
```

```
Query Match    82.3%;  Score 21.4;  DB 2;  Length 164478;
Best Local Similarity   95.7%;  Pred. No. 80;
Matches   22;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
---------------------------- Genome Center ----------------------------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
                     ------ Project Information ------
Center project name: M_BA0285C16
                     ------ Summary Statistics ------
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye primer ET; 0% of reads
Chemistry: Dye terminator big dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180867 bases at least Q40
Consensus quality: 182198 bases at least Q30
Consensus quality: 183076 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 185844; sum-of-contigs
Quality coverage: 7.80 in Q20 bases; agarose-fp
Quality coverage: 6.51 in Q20 bases; sum of contigs
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record
* is arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
  be preserved.
```

FEATURES section (right record):

```
FEATURES             Location/Qualifiers
     source          1..186217
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-285C16"
*                    1..1495
*        1496       1595: contig of 1495 bp in length
*        1596       1595: gap of unknown length
*        1596       3916: contig of 2321 bp in length
*        3917       4016: gap of unknown length
*        4017       7041: contig of 3025 bp in length
*        7042       7141: gap of unknown length
*        7142       9838: contig of 2697 bp in length
*        9839       9938: gap of unknown length
*        9939      16233: contig of 6295 bp in length
*       16234      16333: gap of unknown length
*       16334      23315: contig of 6982 bp in length
*       23316      23415: gap of unknown length
*       23416      34801: contig of 11386 bp in length
*       34802      34901: gap of unknown length
*       34902      51362: contig of 16461 bp in length
*       51363      51462: gap of unknown length
*       51463      73102: contig of 21640 bp in length
*       73103      73202: gap of unknown length
*       73203      98846: contig of 25644 bp in length
*       98847      98946: gap of unknown length
*       98947     117407: contig of 18461 bp in length
*      117408     117507: gap of unknown length
*      117508     154549: contig of 37042 bp in length
*      154550     154649: gap of unknown length
*      154650     185986: contig of 31337 bp in length
*      185987     186086: gap of unknown length
*      186087     186217: contig of 131 bp in length
                     Location/Qualifiers
                     1..186217
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
     misc_feature    1..1495
                     /note="assembly_name:Contig45"
     misc_feature    1596..3916
                     /note="assembly_name:Contig46"
     misc_feature    4017..7041
                     /note="assembly_name:Contig46"
     misc_feature    7142..9838
                     /note="assembly_name:Contig47"
     misc_feature    9939..16233
                     /note="assembly_name:Contig48"
     misc_feature    9939..16233
                     /note="assembly_name:Contig49"
```

RESULT 4

AC069044

LOCUS   AC069044   187592 bp   DNA   linear   HTG 04-MAY-2001
DEFINITION Homo sapiens clone RP11-456A17, WORKING DRAFT SEQUENCE, 29
            unordered pieces.
ACCESSION   AC069044
VERSION   AC069044.2 GI:8649772
KEYWORDS   HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 187592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens clone RP11-456A17
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 187592)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barta,N., Bastien,V., Beda,F.,
            Bonalduk,K.L., Benktai,er,B., Brown,A., Burke,..,
            Campagioano,A., Castle,J., Cheepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,J., DeArellano,K., Devon,K., ...
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Gindo,S., Goyette,M., Graham,J.,
            Grand Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,C., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaBreque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy M., McEwan,P., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mioupa,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,V., Pollara,V., Raymond,C., Riley,R., Rogo,P., Rottman,S.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,H.,
            Stange Thomann,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trujillo,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,D., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jun 24, 2000 this sequence version replaced gi:7884564.
            All repeats were identified using RepeatMasker.
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/cgi-bin/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: 456_A_17
            Center clone name: 456_A_17
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 173381 bases at least Q40
            Consensus quality: 180464 bases at least Q30
            Consensus quality: 181190 bases at least Q20
            Insert size: 180000; agarose-fp
            Insert size: 184892; sum-of-contigs
            Quality coverage: 3.9 in Q20 bases; agarose fp
            Quality coverage: 3.9 in Q20 bases; Sum of contigs
            NOTE: This is a 'working draft' sequence. It currently
            consists of 29 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            this record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *   1   1249: contig of 1218 bp in length
            *   1249   1338: gap of   100 bp
            *   1339   2646: contig of 1408 bp in length
            *   2647   2716: gap of   100 bp
            *   2747   3747: contig of 1001 bp in length
            *   3748   3847: gap of   100 bp
            *   3848   5418: contig of 1571 bp in length
            *   5419   5518: gap of   100 bp
            *   5519   7079: contig of 1561 bp in length
            *   7080   7179: gap of   100 bp
            *   7180   9028: contig of 1849 bp in length
            *   9029   9128: gap of   100 bp
            *   9129   11378: contig of 2250 bp in length
            *   11379   11478: gap of   100 bp
            *   11479   13484: contig of 2006 bp in length
            *   13485   13584: gap of   100 bp
            *   13585   15737: contig of 2153 bp in length
            *   15738   15837: gap of   100 bp
            *   15838   19964: contig of 4127 bp in length
            *   19965   20064: gap of   100 bp
            *   20065   24312: contig of 4248 bp in length
            *   24313   24412: gap of   100 bp
            *   24413   30036: contig of 5624 bp in length
            *   30037   30136: gap of   100 bp
            *   30137   35281: contig of 5145 bp in length
            *   35282   35381: gap of   100 bp
            *   35382   40530: contig of 5149 bp in length
            *   40531   40630: gap of   100 bp
            *   40631   47249: contig of 6619 bp in length
            *   47250   47349: gap of   100 bp
            *   47350   53028: contig of 5679 bp in length
            *   53029   53128: gap of   100 bp
            *   53129   60168: contig of 7040 bp in length
            *   60169   60268: gap of   100 bp
            *   60269   60601: contig of 6344 bp in length
            *   66602   66701: gap of   100 bp
            *   66702   73033: contig of 6332 bp in length
            *   73034   73133: gap of   100 bp
            *   73134   79683: contig of 6550 bp in length
            *   79684   79783: gap of   100 bp
            *   79784   87783: contig of 8000 bp in length

FEATURES   Location/Qualifiers
source   1..187592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-456A17"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature   1..1238
                     /note="assembly_fragment"
misc_feature   1339..2747
                     /note="assembly_fragment"
misc_feature   2747..3747
                     /note="assembly_fragment"
                     /note="assembly_fragment"
                     /clone_end:T7
                     vector_side:left"
misc_feature   3848..5418
                     /note="assembly_fragment"
misc_feature   5519..7079
                     /note="assembly_fragment"
misc_feature   7180..9028
                     /note="assembly_fragment"
misc_feature   9129..11178
                     /note="assembly_fragment"
misc_feature   11479..13484
                     /note="assembly_fragment"
misc_feature   13585..15737
                     /note="assembly_fragment"
misc_feature   15838..19964
                     /note="assembly_fragment"
misc_feature   20065..24312
                     /note="assembly_fragment"
misc_feature   24413..30036
                     /note="assembly_fragment"
misc_feature   30137..35281
                     /note="assembly_fragment"
misc_feature   35382..40530
                     /note="assembly_fragment"
misc_feature   40631..47249
                     /note="assembly_fragment"
misc_feature   47350..53028
                     /note="assembly_fragment"
                     /note="assembly_fragment
                     clone_end:Sp6
                     vector_end:Sp6
                     vector_side:left"
misc_feature   53129..60168
                     /note="assembly_fragment"
misc_feature   60269..6601
                     /note="assembly_fragment"
misc_feature   66702..73033
                     /note="assembly_fragment"
misc_feature   73134..79683
                     /note="assembly_fragment"
misc_feature   79784..8783
                     /note="assembly_fragment"
misc_feature   87884..9500
                     /note="assembly_fragment"
misc_feature   9501..106227
                     /note="assembly_fragment"
misc_feature   10632..115811
                     /note="assembly_fragment"
misc_feature   115912..125733
                     /note="assembly_fragment"

* 87784   87883: gap of   100 bp
* 87884   95000: contig of 7117 bp in length
* 95001   95100: gap of   100 bp
  95101 106227: contig of 11127 bp in length
* 106228 106327: gap of   100 bp
  106328 115811: contig of 9484 bp in length
* 115812 115911: gap of   100 bp
  115912 125833: contig of 9922 bp in length
* 125834 125933: gap of   100 bp
  125934 138211: contig of 12378 bp in length
* 138212 138311: gap of   100 bp
  138312 161514: contig of 23203 bp in length
* 161515 161614: gap of   100 bp
  161615 187592: contig of 25978 bp in length

misc_feature   /note="assembly_fragment"
misc_feature   125834..138211
                     /note="assembly_fragment"
misc_feature   138312..161514
                     /note="assembly_fragment"
misc_feature   161615..187592
                     /note="assembly_fragment"

RESULT 10
LOCUS   AC123369   186653 bp   DNA   linear   HTG 23-JUL-2002
DEFINITION   Rattus norvegicus clone CH230-456P14, *** SEQUENCING IN PROGRESS
             ***, 66 unordered pieces.
ACCESSION   AC123369
VERSION   AC123369.2 GI:21903235
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Rattus norvegicus
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 186653)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
             Harbaria,J., Benton,J., Himage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Cha-ko,J., Chavez,D.,
             Chen,G., Chen,P., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C.D., Coyle,M.D., Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Eddar,D., Edwards,C.C., Elha),C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gundratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
             Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,H.,
             Homsi,F., Howard,S., Huber,T., Hulyk,S., Hume,J., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Liu,J., Liu,J., Liu,W., Lobsiegen,H.,
             Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
             Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
             Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.E., Mel,G., Metzker,M.,
             Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
             Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
             Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
             Oragunye,N., Oviedo,R., Pace,A., Payton,B., Perez,L.,
             Peters,L., Pickens,R., Primus,E., Pu,L.L., Puazo,G., Quiles,M., Ren,Y.,
             Fives,M., Rojas,J., Ruiz,S.J., Saenz,R., Salerno,W., Savery,G.,
             Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
             Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
             Sutton,A., Svatek,A., Svetz,L., Tabor,P., Tamerisa,A., Taylor,C.,
             Lanisby,J., Taylor,C., Taylor,T., Teichroeb,I., Thomas,N., Thomas,H.,
             Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
             Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
             Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
             Wu,C., Wu,Y., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.

```
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 186651)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (29 MAY 2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4  (bases 1 to 186651)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (23-JUN-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jul 18, 2002 this sequence version replaced gi:21290271.
          Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu
          Contact: hgsc-help@bcm.tmc.edu
          ------------------------------------
          Project Information
          Center project name: GWFN
          Center clone name: CH240-122F14
          ------------------------------------
          Summary Statistics
          Sequencing vector: Plasmid;
          Chemistry: Dye terminator Big Dye; 1008 of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 129877 bases at least Q40
          Consensus quality: 137213 bases at least Q30
          Consensus quality: 141956 bases at least Q20
          NOTE: Estimated insert size may differ from sequence length
          (see http://www.ncbi.nlm.nih.gov/Genbank/htgs/draft.html)
          NOTE: This is a 'working draft' sequence. It currently
          consists of 66 contigs. The true order of the pieces
          is not known and their order in this sequence record is
          arbitrary. Gaps between the contigs are represented as
          runs of N, but the exact sizes of the gaps are unknown.
          This record will be updated with the finished sequence
          as soon as it is available and the accession number will
          be preserved.
```

1	1095: contig of 1095 bp in length	
1096	1195: gap of unknown length	
1196	2207: contig of 1012 bp in length	
2209	2397: gap of unknown length	
2398	3492: contig of 1185 bp in length	
2408	4592: gap of unknown length	
4493	4680: contig of 1088 bp in length	
4581	4780: gap of unknown length	
4781	6298: contig of 1518 bp in length	
6299	6498: gap of unknown length	
6499	7792: contig of 1294 bp in length	
7793	7992: gap of unknown length	
7994	9857: contig of 1965 bp in length	
9858	9957: gap of unknown length	
9958	11466: contig of 1409 bp in length	
11467	11566: gap of unknown length	
11567	12793: contig of 1327 bp in length	
12794	12893: gap of unknown length	
12894	14485: contig of 1492 bp in length	
14486	14585: gap of unknown length	
14486	15552: contig of 1067 bp in length	
15553	15652: gap of unknown length	
15653	17523: contig of 1871 bp in length	
17524	17623: gap of unknown length	
17624	19242: contig of 1619 bp in length	
19243	19442: gap of unknown length	
19443	20706: contig of 1164 bp in length	
20707	20806: gap of unknown length	
20807	22585: contig of 1780 bp in length	
22587	22685: gap of unknown length	
22687	24952: contig of 2266 bp in length	
24953	25952: gap of unknown length	

25953	27991: contig of 2049 bp in length	
27992	28091: gap of unknown length	
27192	29146: contig of 1955 bp in length	
29147	29246: gap of unknown length	
29247	30531: contig of 1285 bp in length	
30532	30631: gap of unknown length	
30632	32232: contig of 1601 bp in length	
32233	32332: gap of unknown length	
32333	33819: contig of 1487 bp in length	
33820	33919: gap of unknown length	
33920	36104: contig of 2185 bp in length	
36105	36204: gap of unknown length	
36205	37499: contig of 1295 bp in length	
37500	37599: gap of unknown length	
37600	39486: contig of 1887 bp in length	
39487	39586: gap of unknown length	
39586	41679: contig of 2093 bp in length	
41680	41779: gap of unknown length	
41780	43567: contig of 1788 bp in length	
43568	43667: gap of unknown length	
43668	46253: contig of 2586 bp in length	
46254	46353: gap of unknown length	
46354	46581: contig of 228 bp in length	
46582	46681: gap of unknown length	
46682	48681: contig of 222 bp in length	
48682	50072: gap of unknown length	
50073	51572: contig of 1391 bp in length	
51551	51650: gap of unknown length	
51651	53292: contig of 1442 bp in length	
53293	53392: gap of unknown length	
53393	54985: contig of 1593 bp in length	
54986	55085: gap of unknown length	
55086	56986: contig of 1901 bp in length	
56987	57086: gap of unknown length	
57087	59234: contig of 2148 bp in length	
59235	59334: gap of unknown length	
59335	61169: contig of 1835 bp in length	
61170	61269: gap of unknown length	
61270	63869: contig of 2600 bp in length	
63870	63969: gap of unknown length	
63970	65838: contig of 1869 bp in length	
65839	65938: gap of unknown length	
65939	67741: contig of 1793 bp in length	
67732	67831: gap of unknown length	
67832	70205: contig of 2374 bp in length	
70206	70305: gap of unknown length	
70306	72625: contig of 2320 bp in length	
72626	72725: gap of unknown length	
72726	75701: contig of 3976 bp in length	
75801	76701: gap of unknown length	
76702	79446: contig of 2545 bp in length	
79447	82530: contig of 3084 bp in length	
82531	82630: gap of unknown length	
82631	85482: contig of 2752 bp in length	
85483	85582: gap of unknown length	
85483	87659: contig of 2177 bp in length	
87660	87759: gap of unknown length	
87760	90866: contig of 3107 bp in length	
90867	90966: gap of unknown length	
90967	94513: contig of 3547 bp in length	
94514	94513: gap of unknown length	
94514	96947: contig of 2434 bp in length	
96948	97047: gap of unknown length	
97048	99687: contig of 2640 bp in length	
99688	99787: gap of unknown length	
99788	102905: contig of 3118 bp in length	
102906	103005: gap of unknown length	
103006	106489: contig of 1484 bp in length	
106490	106589: gap of unknown length	
106590	110011: contig of 3422 bp in length	
110012	110111: gap of unknown length	
110112	113307: contig of 3196 bp in length	

```
*   11330R  11407: gap of unknown length
*   11340R  117484: contig of 4077 bp in length

Query Match
  Best Local Similarity   81.5%;
  Matches   23;  Conservative   0;  Mismatches   3;  Indels   0;  Gaps   0;

Qy          1 CCTCCCCAAATCAATTACAATTTCTCC 26
              | ||||||||||||||||||||| |||
Db    17793 CTTTCCAAAATCAATTCAATTTCTCC 17818
```

```
RESULT 11
LOCUS       AC110785
AC110785
LOCUS       AC110785              100987 bp    DNA     linear   PRI 09-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-455L7 from 4, complete sequence.
ACCESSION   AC110785  AC044831
VERSION     AC110785.2  GI:18874242
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100987)
  AUTHORS   Wang,C., Haglund,K. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-455L7
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 100987)
  AUTHORS   Sulston,J.E. and Waterston,R
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   MEDLINE   99067074
   PUBMED    9847074
REFERENCE   2  (bases 1 to 100987)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 100987)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 100987)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 FEB 2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 100987)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 24, 2002 this sequence version replaced gi:18677616.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0455107
            Drafting Center: WIBR
            --------------
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >-
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson. Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11-?? human BAC library was made from the blood of one male
donor as described by Osoegawa,K., Woon,P.Y., Zhao,H., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-507B16, 2000 bp overlap.
The clone sequenced to the right is RP11 435C3, 2000 bp overlap.
Actual start of this clone is at base position 34465 of
RP11-507H10.

Polymorphisms exists between AC110785 and AC020556.

The sequence of AC044831 has been incorporated into AC110785.
FEATURES             Location/Qualifiers
     source          1..100987
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-455L7"
                     /clone_lib="RPCI-11"
     repeat_region   126..873
                     /rpt_family="ERV1"
     repeat_region   1260..1831
                     /rpt_family="ERVL"
     repeat_region   4910..5211
                     /rpt_family="Alu"
     repeat_region   5274..5302
                     /rpt_family="Alu"
     repeat_region   5319..5621
                     /rpt_family="AT_rich"
     repeat_region   5634..5663
                     /rpt_family="Alu"
     repeat_region   5686..5758
                     /rpt_family="AT_rich"
     repeat_region   5997..6368
                     /rpt_family="(TA)n"
     repeat_region   6407..6440
                     /rpt_family="MaLR"
     repeat_region   6575..6598
                     /rpt_family="AT_rich"
     repeat_region   7228..7053
                     /rpt_family="AT_rich"
     repeat_region   7054..7380
                     /rpt_family="AT_rich"
     repeat_region   9177..9632
                     /rpt_family="Alu"
     repeat_region   9937..10192
                     /rpt_family="ERVL"
     repeat_region   10225..10888
                     /rpt_family="L1"
     repeat_region   11315..11433
                     /rpt_family="L2"
     repeat_region   11434..11512
                     /rpt_family="L1"
```

```
repeat_region    1840..1944      /rpt_family "Alu"
repeat_region    1856..14950     /rpt_family="L1"
repeat_region    14068..14144    /rpt_family="L1"
repeat_region    14483..14479    /rpt_family="AT_rich"
repeat_region    14548..14664    /rpt_family="L1"
repeat_region    14699..15082    /rpt_family "L1"
repeat_region    15106..16815    /rpt_family "L1.."
repeat_region    18895..20357    /rpt_family "L1"
repeat_region    20458..20603    /rpt_family="L1"
repeat_region    20604..21627    /rpt_family "Alu"
repeat_region    21216..21924    /rpt_family "AT_rich"
repeat_region    21628..24139    /rpt_family "L1"
repeat_region    21809..22649    /rpt_family "Alu"
repeat_region    21925..24885    /rpt_family "L1"
repeat_region    24574..24923    /rpt_family "Alu"
repeat_region    24781..25082    /rpt_family "L1"
repeat_region    25149..25648    /rpt_family "L1"
repeat_region    25549..25959    /rpt_family "MER1_type"
repeat_region    25651..29448    /rpt_family "Alu"
repeat_region    29093..30109    /rpt_family "MalR"
repeat_region    30010..30224    /rpt_family "L2"
repeat_region    30124..30545    /rpt_family "L2"
repeat_region    30256..32059    /rpt_family "Alu"
repeat_region    31016..32210    /rpt_family "L1"
repeat_region    32164..32876    /rpt_family="L2"
repeat_region    32519..34077    /rpt_family="MalR"
repeat_region    32887..34293    /rpt_family="MaIr"
repeat_region    33850..34247    /rpt_family="MaIr"
repeat_region    34521..34968    /rpt_family "MalR"
repeat_region    34968..45165    /rpt_family "MIR"
repeat_region    45196..45280    /rpt_family="L2"
repeat_region    46154..46374    /rpt_family "L2"
repeat_region    46395..46601    /rpt_family "AU_rich"
repeat_region    46609..37685    /rpt_family "L1"
```

```
Query Match      77.7%;  Score 20.2; DB 9; Length 100987;
Host Local Similarity  88.0%; Pred. No. 2.5e-02;
Matches 22; Conservative  0; Mismatches  1; Indels  0; Gaps  0;
```

```
Qy       1 CCTCCCCAAATCAATTACATTTCT 25
           | |||||||||||||||||||||||
Db   74441 CATCCCCAAATCAATTCCATTTGCT 74465
```

```
RESULT 12
AP004970
LOCUS       AP004970                   106000 bp    DNA        linear
DEFINITION  Lotus japonicus genomic DNA, chromosome 5, clone:LjT40P03,
            complete sequence.
ACCESSION   AP004970
VERSION     AP004970.1  GI:21907988
KEYWORDS    HTG.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S., and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence
            Features of the Genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106000)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:; kazusa.or.jp,
            URL:http://www.kazusa.or.jp; Tel:81-438-52-3935;
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..106000
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="5"
                     /clone="LjT40P03"
                     /note="LjT library"
                     /clone_lib="LjT library"
                     /note="TAC clone:LjT40P03"
BASE COUNT    33621 a  19351 c  18804 g  14224 t
ORIGIN
```

```
Qy       1 CCTCCCCAAATCAATTACATTTCT 25
           |||| ||||||||| | |||||||
Db   77429 CCTCCCAAATCAATTGCCTTTTC 77454
```

```
RESULT 13
AC073450/c
LOCUS       AC073450                 167623 bp    DNA        linear
DEFINITION  Homo sapiens chromosome 4 clone RP11-521C6 map 4, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC073450
VERSION     AC073450.2  GI:10567951
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167623)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-521C6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167623)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
```

TITLE   Direct Submission
JOURNAL Submitted (18-JUN-2000) Whitehead Institute/MIT Center for Genome
         Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Oct. 4, 2000 this sequence version replaced gi:8571742.
         All repeats were identified using RepeatMasker;
         Smit, A.F.A. & Green, P. (1996-1997)
         http://ftp.genome.washington.edu/RM/RepeatMasker.html
         ----
         Center: Whitehead Institute/ MIT Center for Genome Research
         Center code: WIBR
         Web site: http://www-seq.wi.mit.edu
         Contact: sequence_submissions@genome.wi.mit.edu

         Project information
         Center project name: LL0444
         Center clone name: 521_C_6

         Summary Statistics
         Sequencing vector: M13; M77815; 100% of reads
         Chemistry: Dye terminator Big Dye; 100% of reads
         Assembly program: Phrap; version 0.960731
         Consensus quality: 157250 bases at least Q40
         Consensus quality: 162563 bases at least Q30
         Consensus quality: 164497 bases at least Q20
         Insert size: 167623; sum-of-contigs
         Quality coverage: 4.3 in Q20 bases; sum-of-contigs
         ----
       * NOTE: This is a 'working draft' sequence. It currently
       * consists of 21 contigs. The true order of the pieces
       * is not known and their order in this sequence record is
       * arbitrary. Gaps between the contigs are represented as
       * runs of N, but the exact sizes of the gaps are unknown.
       * This record will be updated with the finished sequence
       * as soon as it is available and the accession number will
       * be preserved.
       *
       *   1   586: contig of 586 bp in length
       *   587   686: gap of 100 bp
       *   687   1860: contig of 1174 bp in length
       *   1861   1960: gap of 100 bp
       *   1961   3559: contig of 1599 bp in length
       *   3560   3659: gap of 100 bp
       *   3660 31851: contig of 28192 bp in length
       *   31852 31951: gap of 100 bp
       *   31952 33013: contig of 1062 bp in length
       *   33014 33113: gap of 100 bp
       *   33114 35199: contig of 2086 bp in length
       *   35200 35299: gap of 100 bp
       *   35300 38384: contig of 3085 bp in length
       *   38385 38484: gap of 100 bp
       *   38485 43409: contig of 4925 bp in length
       *   43410 43509: gap of 100 bp
       *   43510 48273: contig of 4764 bp in length
       *   48274 48373: gap of 100 bp
       *   48374 53158: contig of 4785 bp in length
       *   53159 53258: gap of 100 bp
       *   53259 60666: contig of 7408 bp in length
       *   60667 60766: gap of 100 bp

Campopiano,A., Castle,A., Chapel,L., Chiangelo,M., Collins,S.,
Collymore,D., Cook,D., DeArellano,K., Dewar,K., Diaz,S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,F., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Kleir,J., Laperty,K., Lamarche,L., Landers,T., Lehoczky,J.,
Levine,R., Liew,C., Liu,C., Locke,K., Macdonald,J., Marquis,N.,
McCarthy,M., McEwan,P., McKrth,A., McKnian,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nusbaum,C., O'Donnell,T.,
O'Neil,P., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,P., Ro, y,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,T.,
Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

FEATURES   Location/Qualifiers
     source   1..167623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-521C6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..586
                     /note="assembly_fragment"
                     /vector_side:left
     misc_feature   687..1860
                     /note="assembly_fragment"
     misc_feature   1961..3559
                     /note="assembly_fragment"
     misc_feature   3660..31851
                     /note="assembly_fragment"
     misc_feature   31952..33013
                     /note="assembly_fragment"
     misc_feature   33114..35199
                     /note="assembly_fragment"
     misc_feature   35300..38384
                     /note="assembly_fragment"
     misc_feature   38485..43409
                     /note="assembly_fragment"
     misc_feature   43510..48273
                     /note="assembly_fragment"
     misc_feature   48374..53158
                     /note="assembly_fragment"
     misc_feature   53259..60666
                     /note="assembly_fragment"
     misc_feature   60767..68003
                     /note="assembly_fragment"
     misc_feature   68104..77478
                     /note="assembly_fragment"
     misc_feature   77579..86611
                     /note="assembly_fragment"
     misc_feature   86712..95048
                     /note="assembly_fragment"
     misc_feature   95149..107974
                     /note="assembly_fragment"
     misc_feature   108075..122534
                     /note="assembly_fragment"
     misc_feature   122635..137590
                     /note="assembly_fragment"
     misc_feature   137691..152722
                     /note="assembly_fragment"
     misc_feature   152823..167352
                     /note="assembly_fragment"
     misc_feature   167453..167623
                     /note="assembly_fragment"
                     /vector_side:right

       *   60767 68003: contig of 7237 bp in length
       *   68004 68103: gap of 100 bp
       *   68104 77478: contig of 9375 bp in length
       *   77479 77578: gap of 100 bp
       *   77579 86611: contig of 9033 bp in length
       *   86612 86711: gap of 100 bp
       *   86712 95048: contig of 8337 bp in length
       *   95049 95148: gap of 100 bp
       *   95149 107974: contig of 12826 bp in length
       *   107975 108074: gap of 100 bp
       *   108075 122534: contig of 14460 bp in length
       *   122535 122634: gap of 100 bp
       *   122635 137590: contig of 14956 bp in length
       *   137591 137690: gap of 100 bp
       *   137691 152722: contig of 15032 bp in length
       *   152723 152822: gap of 100 bp
       *   152823 167352: contig of 14530 bp in length
       *   167353 167452: gap of 100 bp
       *   167453 167623: contig of 171 bp in length

BASE COUNT
ORIGIN

RESULT 14
AC107H04/c
LOCUS   AC107H04
DEFINITION Mus musculus clone RP24 48C16, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC107H04
VERSION   AC107H04.4 GI:22981958
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_POLBOP.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 240419)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Mus musculus, clone RP24-48C16
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 240419)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,M., Bastien,V., Bogdslavski,I., Boukhgalter,P
            Brown,A., Camarata,J., Campgrzano,A., Chang,J., Chazaro,B.,
            Cheepel,Y., Colando,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., DeWitt,K., Dietz,S., Dodge,S., Fare,S.,
            Ferriera,P., FitzHugh,W., Saxe,D., Calagan,J., Gardyna,S.,
            Gnde,S., Gurd,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hayes,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Kalus,A., Kells,C., LaRoque,K., Lamarares,R.,
            Landers,T., Lebowitzky,J., Levine,R., Lin,D., MacLean,C.,
            Macdonald,P., Major,J., Marquis,P., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Mendos,T., Mihova,T.,
            Mienta,V., Murphy,T., Naylor,J., Nguyen,T., Nicol,R., Narbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,M., Phuaboung,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rioback,M., Riley,R., Rise,C., Rogov,P., Roman,A.,
            Rosette,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,D., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,J., Subramanian,A., Talamas,J., TesLaye,S., Theodore,J.,
            Topham,K., Travis,M., Travis,N., Trigillio,J., Vassiliev,H.,
            Viel,K., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 240419)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,M., Bastien,V., Bloom,T., Bogdslavski,I., Houkhgalter,B.,
            Camarata,J., Chang,J., Chazaro,H., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., DewItt,K., Dietz,S., Dodge,S.,
            Fare,S., Ferriera,P., FitzGerald,M., Gage,D., Galagan,J.,
            Girjha,S., Gnde,S., Graham,L., Grand-Pierre,N., Hayes,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamal,A.,
            Katras,A., Kells,C., Landers,P., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., McEwan,P., Meldrim,J., Mendos,T., Meta,V.,
            Murphy,P., Nayltr,J., Nguyen,C., Nicol,R., Narbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phuaboung,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,D., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Testaye,S., Theodore,J., Topham,K., Travis,M., Vassiliev,H.,
            Viel,K., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.

COMMENT   Direct Submission
JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 21, 2002 this sequence version replaced gi:21426296.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996,1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center --------------------
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIHR
            Web site: http://www.seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------- Project Information -----------
            Center project name: L20527
            Center clone name: 48_C_16
            --------- Summary Statistics ---------
            Sequencing Vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960741
            Consensus quality: 229651 bases at least Q40
            Consensus quality: 229770 bases at least Q30
            Consensus quality: 229770 bases at least Q20
            Insert size: 225000; agarose-fp
            Insert size: 229019; sum-of-contigs
            Quality coverage: 11.8 in Q40 bases; agarose-fp
            Quality coverage: 11.6 in Q20 bases, sum of contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1 240419
            * Location/Qualifiers
FEATURES
source   1..240419
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP24-48C16"
                /clone_lib="RPCI 23 Female Mouse BAC"
misc_feature   1..1429
                /note
                /clone_end=SP6
                /vector_side=left"
misc_feature   1510..22315
                /note
                /clone_end=T7
                /vector_side=right"
misc_feature   22416..77795
                /note="assembly_fragment"
misc_feature   77896..125553
                /note="assembly_fragment"
misc_feature   125654..188794
                /note="assembly_fragment"
misc_feature   188895..230419
                /note="assembly_fragment"

            *   1429: gap of   100 bp
            *   1430: contig of 1429 bp in length
            *   1510: gap of   20786 bp in length
            *   22315: gap of   100 bp
            *   22316: contig of 20786 bp in length
            *   22415: contig of 55480 bp in length
            *   77795: gap of   100 bp
            *   77896: contig of 47658 bp in length
            *   125553: gap of   100 bp
            *   125654: contig of 63141 bp in length
            *   188794: gap of   100 bp
            *   188895: contig of 41525 bp in length
            *   230419: contig of 41525 bp in length

BASE COUNT   59296 a 4405 c 4435 g 74208 t 545 others
ORIGIN

Query Match   77.7%; Score 20.2; DB 2; Length 240419;

```
Best Local Similarity 89.0%, predicted No.2 loop02:
Matches 22, Conservative 0, Mismatches 3, Indels 0, Gaps 0;

QY       1 CTCTCCCAAATCAAATTAGATTTCT 25
           ||||||||||||||||||||  |||
Db 165061 CTACCCCAAATAAAATTAGACTTTCCT 165037
```

RESULT 15
AC016175/c
LOCUS   AC016175   95445 bp   DNA   linear   HTG 13-JUL-2000
DEFINITION   Homo sapiens clone RP11-20M11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016175
VERSION   AC016175.2 GI:9123967
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE   Homo sapiens
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 95445)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-20M11
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 95445)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P, DeArellano,K., Dewar,K., Dominic,M., Donelan,I., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
             Calagan,J., Gardyna,S., Grant,G., Hafez,B., Hadford,A., Horton,L.,
             Howland,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lui,A., Lokyitsang,T., Macdonald,P., Marquis,N.,
             McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
             Morrow,T., Naylor,J., Nichahni,H., O'Connell,T., O'Donnell,P.,
             Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,F., Sciarty,P.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Testaye,S., Tirolli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
             Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 13 2000 this sequence version replaced gi:6466619.
             All repeats were identified using RepeatMasker
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ----
             Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu

             -- Project Information --
             Center project name: L3991
             Center clone name: 20_M_11

             NOTE: This record contains 91 individual
             sequencing reads that have not been assembled into
             contigs. Runs of N are used to separate the reads
             and the order in which they appear is completely
             arbitrary. Low pass sequence sampling is useful for
             identifying clones that may be gene-rich and allows
             overlap relationships among clones to be deduced.
             However, it should not be assumed that this clone
             will be sequenced to completion. In the event that
             the record is updated, the accession number will
             be preserved.

             *   1 1043: contig of 943 bp in length
             *   944 1043: gap of 100 bp
             *   1044 2014: contig of 971 bp in length
             *   2015 2114. gap of 100 bp
             *   2115 3045: contig of 931 bp in length
             *   3046 3145: gap of 100 bp
             *   3146 4117: contig of 992 bp in length
             *   4118 4237: gap of 100 bp
```

```
*   4238   5204: contig of  957 bp in length
*   5205   5304: gap of  100 bp
*   5305   6255: contig of  951 bp in length
*   6256   6355: gap of  100 bp
*   6356   7315: contig of  960 bp in length
*   7316   7415: gap of  100 bp
*   7416   8363: contig of  948 bp in length
*   8364   8463: gap of  100 bp
*   8464   9416: contig of  953 bp in length
*   9417   9516: gap of  100 bp
*   9517  10473: contig of  957 bp in length
*  10474  10573: gap of  100 bp
*  10574  11524: contig of  951 bp in length
*  11525  11624: gap of  100 bp
*  11625  12565: contig of  941 bp in length
*  12566  12665: gap of  100 bp
*  12666  13609: contig of  944 bp in length
*  13610  13709: gap of  100 bp
*  13710  14666: contig of  957 bp in length
*  14667  14766: gap of  100 bp
*  14767  15743: contig of  977 bp in length
*  15744  15843: gap of  100 bp
*  15844  16791: contig of  948 bp in length
*  16792  16891: gap of  100 bp
*  16892  17843: contig of  952 bp in length
*  17844  17943: gap of  100 bp
*  17944  18920: contig of  977 bp in length
*  18921  19020: gap of  100 bp
*  19021  19953: contig of  933 bp in length
*  19954  20053: gap of  100 bp
*  20054  20979: contig of  926 bp in length
*  20980  21079: gap of  100 bp
*  21079  22015: contig of  937 bp in length
*  22016  22115: gap of  100 bp
*  22115  22857: contig of  743 bp in length
*  22857  23065: contig of  100 bp in length
*  23065  24107: contig of  942 bp in length
*  24107  24207: gap of  100 bp
*  24208  25126: contig of  919 bp in length
*  25127  25226: gap of  100 bp
*  25227  26185: contig of  960 bp in length
*  26286  26287: gap of  100 bp
*  26287  27240: contig of  954 bp in length
*  27241  27340: gap of  100 bp
*  27341  28509: contig of  969 bp in length
*  28510  28609: gap of  100 bp
*  28510  29463: contig of  955 bp in length
*  29464  29563: gap of  100 bp
*  29564  30435: contig of  371 bp in length
*  30435  30535: gap of  100 bp
*  30535  31476: contig of  941 bp in length
*  31477  31576: gap of  100 bp
*  31577  32532: contig of  956 bp in length
*  32533  32632: gap of  100 bp
*  32632  33591: contig of  959 bp in length
*  33592  33691: gap of  100 bp
*  33692  34636: contig of  945 bp in length
*  34637  34736: gap of  100 bp
*  34737  35689: contig of  953 bp in length
*  35690  35789: gap of  100 bp
*  35790  36751: contig of  962 bp in length
*  36752  36851: gap of  100 bp
*  36852  37789: contig of  938 bp in length
*  37790  37889: gap of  100 bp
*  37890  38859: contig of  970 bp in length
*  38860  38959: gap of  100 bp
*  38960  39864: contig of  905 bp in length
*  39865  39964: gap of  100 bp
*  39965  40924: contig of  960 bp in length
*  40925  41024: gap of  100 bp
*  41025  41990: contig of  966 bp in length
*  41991  42090: gap of  100 bp
*  42091  43002: contig of  912 bp in length
```

```
•   44003   44102: gap of    100 bp
•   44103   44051: gap of    100 bp
•   44052   44151: gap of    100 bp
•   44152   45089: contig of  918 bp in length
•   45090   45189: gap of    100 bp
•   45190   46145: contig of  946 bp in length
•   46146   46245: gap of    100 bp
•   46246   47158: contig of  934 bp in length
•   47169   47268: gap of    100 bp
•   47269   48208: contig of  940 bp in length
•   48209   48408: contig of  100 bp in length
•   48409   49276: contig of  968 bp in length
•   49277   50334: contig of  958 bp in length
•   45177   50434: gap of    100 bp
•   50435   51116: contig of  882 bp in length
•   51117   51416: gap of    100 bp
•   51417   52392: contig of  976 bp in length
•   52393   52492: gap of    100 bp
•   52493   53398: contig of  906 bp in length
•   53499   53498: gap of    100 bp
•   53499   54459: contig of  961 bp in length
•   54460   54559: gap of    100 bp
•   54559   55516: contig of  957 bp in length
•   54560   55516: contig of  959 bp in length
•   55517   55575: gap of    100 bp
•   55576   56576: contig of  964 bp in length
•   56576   57748: gap of    100 bp
•   57749   58700: contig of  968 bp in length
•   58707   58806: gap of    100 bp
•   58807   59747: contig of  931 bp in length
•   59748   59837: gap of    100 bp
•   59838   60776: contig of  939 bp in length
•   60777   60876: gap of    100 bp
•   60877   61862: contig of  986 bp in length
•   61863   61962: gap of    100 bp
•   61963   62954: contig of  992 bp in length
•   62955   63054: gap of    100 bp
•   63055   64051: contig of  897 bp in length
•   64052   64951: gap of    100 bp
•   64952   65003: contig of  952 bp in length
•   65004   65103: gap of    100 bp
•   65104   66057: contig of  954 bp in length
•   66058   66157: gap of    100 bp
•   66158   67112: contig of  955 bp in length
•   67113   67212: gap of    100 bp
•   67213   68176: contig of  964 bp in length
•   68177   68276: gap of    100 bp
•   68277   69207: contig of  931 bp in length
•   69208   69307: gap of    100 bp
•   69308   70204: contig of  896 bp in length
•   70204   70304: gap of    100 bp
•   70404   71245: contig of  942 bp in length
•   71246   71345: gap of    100 bp
•   71346   72289: contig of  954 bp in length
•   72290   72389: gap of    100 bp
•   72390   73363: contig of  974 bp in length
•   73464   73463: contig of  100 bp in length
•   73464   74456: contig of  993 bp in length
•   74457   74556: gap of    100 bp
•   74557   75492: contig of  946 bp in length
•   75494   75592: contig of  100 bp
•   75593   76525: contig of  933 bp in length
```

Query Match
Best Local Similarity   76.28;   Score 19.8;   OH 2;   Length 95445;
Matches 21;   Conservative 0;   Pred. No. 3.8e+02;   Mismatches 2;   Indels 0;   Gaps 0;

```
Qy    3  TCCCCAAATCAATTAGCATTTCT  25
         ||||||||||||||||| || || ||
Db  6484  TCCCCAAATCAATTACTTTTCT  6462
```